BLANK

11 A Rational Analysis of Human Memory

John R. Anderson
Carnegie-Mellon University

My approach to human memory has been to search for mechanisms to explain the observed phenomena; this has probably been the dominant approach in the field. It has had its notable successes but also its notable failures. High among the list of such failures has been the inability to resolve the many theoretical dichotomies, such as the status of short-term memory (Crowder, 1982; Wickelgren, 1973), parallel versus serial processing (Townsend, 1974), semantic versus episodic memory (Tulving, 1983), and imaginal versus propositional representations (Anderson, 1978; Kosslyn, 1980; Pylyshyn, 1981). One gets the impression that mechanistic theories of memory involve a precision in theoretical specification that cannot be supported by behavioral data. Without repudiating the mechanistic approach, this chapter is devoted to exploring an alternative way of casting a theory of memory. This alternative formulation starts with the following principle about human memory:

> *Principle of Rationality:* Human memory behaves as an optimal solution to the information-retrieval problems facing humans.

It is called a principle of rationality and it is like the rational man hypothesis in economics. Just as economic behavior is supposedly predictable on the assumption that humans are rational and act to optimize their economic interests, so this principle asserts that human memory is rationally designed to produce optimal performance in human information-retrieval tasks. But how could one entertain for more than a moment the proposition that memory behaves optimally? Aren't we always failing to retrieve the memory we want? Taking long periods to do retrievals that computers do instantaneously? Misremembering what it was that

we experienced? If memory were rationally designed, surely it would deliver to us instantaneously and exactly the memories we want. However, such anti-rationality arguments make two unreasonable assumptions:

1. Memory can know what fact we want before memory has retrieved and tested that fact. This assumption is much too strong. The cues that the environment provides for us are only probabilistic. Until we retrieve and carefully test a fact, we cannot determine if it is relevant to the current needs. The best that one can assume is that memory can make some best guesses as to what facts out of our great storehouse we want and offer these first for judgment.

2. Things can be done instantaneously. This assumption is clearly wrong—there must be some time costs to the acts of memory. In particular, there has to be a cost associated with testing a retrieved memory and determining if it is relevant.

Basically, the point is that memory cannot be omniscient or omnipotent. The principle of rationality only requires that it be optimal.

To apply the principle of rationality to human memory, it is necessary to have some way of framing the information-retrieval problem faced by humans that embodies the two preceding constraints. Framing the optimization problem is where the real theory lies in a rational analysis. For instance, much of the current debate about the rational analysis in economics is not whether human economic behavior is rational, but rather how to frame human economic behavior so that it will be rational. In proposing a rational analysis of human memory, it seems reasonable to start with a framing of the information-retrieval problem that is as simple and bland as possible. Only if that proves inadequate should we go to a more complex framing.

AN ANALYSIS OF INFORMATION RETRIEVAL

There already exists an analysis of information retrieval in the subfield of computer science called, curiously enough, "information retrieval" (Salton & McGill, 1983). The generic information-retrieval system has a data base of stored items and must respond with an ordered subset of these given a query that consists of some key words. Perhaps our most frequent use of such systems in academia is in library searches where we provide some content words and the system responds with a list of possible books and their abstracts. Like human memory, computer information-retrieval systems cannot know what the user really wants. They can only make wise guesses. Secondly, there are real costs in a system associated with mistaken guesses. As with human memory, the system may fail to retrieve the desired items, which clearly is a costly error. However,

there is also a cost associated with retrieving an inappropriate item—which is the user's cost in considering it and rejecting it. Thus, the information-retrieval system cannot just deal with the problem of undergeneration by retrieving everything. In the field of information retrieval, the problem of generating the desired items is called *recall* and the problem of not generating irrelevant items is called *precision*. We can now specify the basic information-retrieval problem for computers—given a query, provide an ordered list of items that provides a maximal combination of recall and precision. I postpone defining "maximal combination" until I come to considering the human situation.

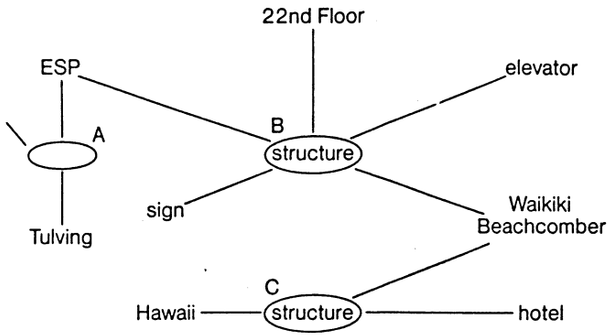
This is the information-retrieval problem as it is conventionally conceived of in computer science. Now let's consider how this maps onto human information retrieval:

1. The items to be retrieved are units of human memory. I refer to these as *structures*, neglecting the issue of whether these units are propositions, productions, images, associations, schemata, or whatever your favorite flavor of cognitive chunk. The important feature of a structure is that it consists of a number of *terms*. Figure 11.1 illustrates some structures that I carry around in my memory. I have a memory of stepping out of the elevator on the 22nd floor of the Waikiki Beachcomber and seeing a sign announcing the ESP floor. There is a memory structure, B, in Fig. 11.1 encoding this memory with terms "22nd floor," "elevator," "Waikiki Beachcomber," "ESP," "sign," and possibly others. Also shown in Fig. 11.1 are memory structures A, relating the terms Tulving and ESP, and C, relating Waikiki Beachcomber, Hawaii, and hotel. It is a further assumption that all the items that were ever recorded as memory structures are still there to be retrieved. Whether this assumption is literally true, or human memory just behaves as if it were, is irrelevant to this rational analysis.

2. The query that prompts our memory is a set of terms that we encode from the current context. Thus, if I am asked, "How does Endel Tulving remind you of Hawaii?", among the terms that would be part of my current memory query would be, "Endel Tulving" and "Hawaii." However, the total query set consists of anything attended to. So it would presumably also include the person asking the question and perhaps other contextual elements.

3. At any point in time, there is a subset of my memory structures that are the targets. They are structures that will help me answer my current information-retrieval demands. For instance, given the example query in the preceding paragraph (2.), a sufficient set of structures would be A, B, and C. In an explicit memory test, the information-processing demands being placed on us may seem pretty clear, but presumably, all throughout life we have to retrieve information to respond to the demands of the current situation.

4. There is a cost associated with failing to retrieve the necessary information. There is also a cost associated with our system retrieving an irrelevant item,



Query

How does [Endel Tulving] remind you of [Hawaii] ?

Targets

Structures sufficient to allow you to calculate an answer to the current problem — A, B & C above.

FIG. 11.1. Files and terms.

which we then have to process and reject (by "process," nothing is implied about whether that processing is available to consciousness). Let us denote the cost associated with a failed recall as C_R for recall failure cost, and the cost associated with processing an irrelevant item as C_P for lack of precision cost. These costs are somewhat analogous to Type 1 and Type 2 errors in statistics. The nature of the recall failure cost should be apparent. I think of the lack of precision cost as largely coming from the wasted effort in testing and rejecting irrelevant memories.

Now we are in the position to define the optimization problem for human memory. Let p[A] be an estimated probability that memory structure A is relevant—that is, a target. A rationally designed information-retrieval system would retrieve memory structures ordered by their probabilities p[A] and stop retrieving when a structure A was encountered such that:

$$p[A]C_R < (1-p[A])C_P \quad (1)$$

That is, the system should stop retrieving when the probabilities are so low that the expected loss due to failing to recall a target item is less than the expected

loss due to retrieving an irrelevant item. Presumably, in most situations, C_R is much larger than C_P so that this inequality will only hold for very small p[A], but considering the number of structures in memory, most would have very small probabilities of being relevant. It should also be noted that in a particular situation, 0, 1, or more memory structures might be relevant. Thus, the p[A] are not constrained to sum to 1. To be able to predict speed and accuracy of recall, we need to inquire as to what factors memory can use to estimate p[A], and our prediction will be that these factors determine memory performance.

This discussion of the optimization problem is framed in serial and conscious terms—first the subject considers one target structure, then another, and so on. However, clearly, the subject need not be aware of the consideration of memory structures. It should also be clear, given our knowledge of the parallel-serial equivalence (Townsend, 1974), that the serial terminology is nothing more than an expository convenience. Indeed, I think of this all as being implemented in the parallel pattern-matching machinery of ACT* (Anderson, 1983). In ACT*, the system can assign resources to the structures it is processing according to their plausibility, and the system can effectively ignore structures below some threshold of plausibility. Thus, whether parallel or serial, the critical feature is that knowledge structures are consulted in order of plausibility until they become too implausible to consider. It is not the goal of this chapter to inquire as to the mechanisms that achieve this ordering, only to inquire whether we can predict memory performance assuming that memory does achieve this ordering.

ESTIMATION OF LIKELIHOODSOF MEMORY STRUCTURES

We are now one big step from having a theory that specifies the behavior of memory from purely rational considerations. That one big step is to specify the p[A] in the preceding discussion.

One solution to the estimation of p[A] that appears in the computer information retrieval literature (Bookstein & Swanson, 1974, 1975) is to use Bayesian estimation procedures. The two obvious pieces of information for evaluating whether a memory structure will be relevant is its past history and the terms in the query. Thus, each structure A has some history H_A of being relevant in the past. The current context consists of a set of terms that I call *cues* and denote by indices i. I denote the set of cues as Q, for query. In doing Bayesian estimation, we are trying to calculate the posteriori probabilities, giving us the following equation:

$$\frac{p(A|H_A\&Q)}{p(A|HA\&Q)} = \frac{p(A|H_A)}{p(A|H_A)} \times \prod_{i\in Q} \frac{p(i|A)}{p(i|A)} \quad (2)$$

That is, the odds ratio for item A is the product of the odds ratio for item A given history H_A times the product of the ratios of the conditional probabilities for each cue i in the context. This equation assumes that all the cues and history are independent. This is a strong assumption, but one that is typically made in the computer information-retrieval literature for purposes of tractability.[1] The first item, $p(A \mid H_A)/p(\bar{A} \mid H_A)$, is basically a prior odds ratio for the item, given its frequency and recency of occurrence. The other quantities, the $p(i \mid A)/p(i \mid \bar{A})$, are the odds ratios of the conditional probabilities of the cues, given that the structure is relevant versus not relevant. These ratios can be thought of as associative strengths. One can imagine a great matrix giving these values for each combination of cue and memory structure. Indeed, if we were to listen to connectionists, this is just what we would believe is in the head. However, these implementation details are not really relevant to the goal of this chapter. Our concern is with what factors should rationally influence the cue strengths and the prior history factor.

The basic behavioral assumption is that memory performance is monotonically related to the preceding ratio, which will be called structure A's likelihood ratio. It would be possible to develop a mapping of likelihood ratio into probability of recall and latency measures in memory research. However, this would be unnecessary detail for the purposes of this chapter. The assumption of a monotonic relationship is good enough to make the points that we are concerned with. The purposes of this chapter are to determine the factors that rationally should determine the history factor or the cue strengths, determine the qualitative predictions of these factors, and ascertain whether these predictions are in fact correct. I deal only with the most basic memory effects—all of which embody ordinal relationships only. It would be of interest to try to map likelihood into probability of recall and latency and see if we can reproduce exact numbers, but that is another paper.

THE HISTORY FACTOR

To address the history factor, we need to determine how the past history of a structure's usage predicts its current usage. To determine this in the most valid way, we would have to follow people about in their daily lives, keeping a complete record of when they use various facts. Such an objective study of human information use is close to impossible. What is possible is to look at records from nonhuman information-retrieval systems that can be objectively studied. For instance, such studies have been done of borrowing from libraries (Burrell, 1980; Burrell & Cane, 1982) and access to computer files (Stritter, 1977). Both systems tend to yield rather similar statistics. If we believe that the

[1]Human memory may not be so constrained and it is interesting to inquire as to which of our predictions might be upset by nonindependence.

statistics of human memory information retrieval mirror the statistics of these nonhuman systems, we are in a position to make predictions about how the human should estimate need probabilities, given past history.

Burrell has developed a mathematical theory of usage for information-retrieval systems such as libraries (a very similar model appears in Stritter, 1977 for file usage). His theory involves two layers of assumptions. First, Burrell assumes that the items (books, files, memory structures) in an information retrieval system vary in terms of their desirability. He assumes that they vary as a gamma distribution with parameter b and index v (i.e., if $v=1$ this is an exponential with parameter b). Such a distribution will produce mean desirability values of v/b and variances v/b^2. Second, given an item with desirability λ the time to next use is an exponential process with mean $1/\lambda$.

If we assume that this model describes the statistics of usage in the human memory system, we are in position to calculate the historical component of the need function. Suppose we observe a particular memory structure that was introduced t time units ago and has been used n times since its introduction. What we have to do in calculating the need function is to determine the probability that this item will be used in the next time unit. We denote this probability by $RF(t,n)$ and it is given by the following formula:

$$RF(t,n) = 1 - \left[\frac{t+b}{t+b+1}\right] \tag{3}$$

If $t + b$ is much larger (eg., 10 times) than $v + n$, Equation 3 is approximated by the following expression:

$$RF(t,n,T) \approx \frac{v+n}{t+b} \tag{3-1}$$

We can always guarantee this approximation by using small enough time units. Or said otherwise, if the interval that we are predicting for is small enough, this approximation will be close.

Equation 3 can be simplified in various ways to address some of the basic memory variables. Let us consider the retention function where we wait t seconds after an item was studied and first test it. In this case, n is zero. The equation now takes the form $v/(t + b)$. A function of this form would give a very good mimicry of human retention functions that are typically described by power functions. Newell & Rosenbloom (1981) concluded that it was basically impossible to distinguish hyperbolic functions such as the one implied by Burrell's model from power functions. If we assume that the human information-retrieval system faces a distribution of demands similar to the objectively observable distributions for retrieval systems like libraries or file systems, then the principle of rationality would predict the observed human retention function. This is the first success for the rational analysis.

We can also examine what this function implies about the effect of n, which is the number of times the structure has been practiced holding t constant. The

reader will observe that it implies that need probability has a linear relationship to practice with a positive, non-zero intercept. (Because we are working with small time intervals and low need probabilities, we don't have to worry about non-linearity as the probability approaches one). What this implies about the learning function depends on how need probability is mapped into observable measures. It turns out, for instance, that ACT* (Anderson, 1983) assumed a linear growth in strength of just this form and did a nice job of predicting learning curves. The ability to predict the learning curve is the second success for the rational analysis.

According to Burrell's model, it does not matter what the spacing of these n presentations is. All that matters in Burrell's model is the total number of presentations (n) and the total elapsed time (t). This lack of sensitivity to spacing is a consequence of the ahistorical character of the exponential process that characterizes the intervals between uses of an item. The question arises whether Burrell's model correctly describes the likelihood function. Is it the case that in information retrieval systems there is no massing of need? Burrell's model implies that the n presentations should be uniformly spaced over the t interval. In fact, Burrell's model is not descriptively accurate here, as one might expect. For instance, in Carnegie-Mellon's library system, there are very definite clusterings of borrowings and one can reject the hypothesis of uniform distribution of the borrowings of a book over a fixed time interval. There are lots of reasons for such massings, such as a book being relevant to a course taught in only one semester. Stritter (1977) noted such deviations from uniformity, but chose to ignore them in developing his model of file systems. It is fairly intuitive that the same is true of human memory, although it is hard to verify the human likelihood function objectively.

If some use is massed and some is not, then the intervals between successive uses should predict the probability of the item being needed now. Thus, compare one item that has been used fairly uniformly n times over the year and another item whose n uses all occurred in a 3-month period 6 months ago. Clearly, the first is more likely to be needed now. We would therefore predict better memory for spaced items, as long as we are not comparing to a massing of study that has just occurred. It seems then that the spacing effect is a third success for the rationality analysis. However, we have no formal model for the clustering that does occur in information retrievals. Thus, we can do more than make the prediction that after sufficient delay, human memory should show better performance for those items whose studies are widely distributed.

CUE STRENGTH

The preceding analysis has been concerned with analyzing the history factor that was the first term in Equation 2. Now we turn to calculating the remaining quantities, $p(i \mid A)/p(i \mid \bar{A})$, which are the cue strengths. We assume that the

way the system assesses these probabilities is by comparing the cue terms against the terms that occur in the structure. Thus, the system might compare the cue term "Tulving" against the memory term "ESP." Intuitively, the cue strengths should increase as the cue terms are more related to the terms in the memory trace. The formal definition of relatedness that I use is taken whole cloth out of the text on information retrieval by Salton & McGill (1983).

Basically, two terms are regarded as related to the degree that they appear in the same memory structures. Let t_{ik} be defined as 1 if term i occurs in memory structure k, and 0 otherwise. Then we can define the relatedness between terms i and j as follows:

$$r_{ij} = \frac{\sum_k t_{i,k}t_{j,k}}{\sum_k t_{i,k} + \sum_k t_{j,k} - \sum_k t_{i,k}t_{j,k}} \tag{4}$$

where the sums are over all structures in memory. The basic intuition is that if i and j are highly related, one is likely to want a structure involving j if one sees i, or vice versa. However, it will take a fair bit of mathematics to relate this basic intuition to our Bayesian framework[2]. Let A be a structure containing e terms. We define the average relatedness between A and i as:

$$R[A,i] = (\sum_j r_{ij})/e \tag{5}$$

where the sum is over the terms in structure A and e is the number of such terms. We can define the proportion association between A and i, or normalized relatedness, as

$$N[A,i] = R[A,i]/\,(\sum_j R[A,j]) \tag{6}$$

where the sum is over all possible terms. The final quantity we need to know is what the association is between structures other than A and i. It is given by the following formula:

$$N[\bar{A},i] = \sum_{X \neq A} N(X,i)/(M-1) \tag{7}$$

where M is the total number of structures.

These quantities figure in the calculation of the cue strengths, $P(i \mid A)/P(i \mid \bar{A})$. We first focus on calculating $P(i \mid A)$. The basic assumption is that the probability of a cue being present is a function of the normalized relatedness of that cue to

[2]The mathematics also strikes me as a bit arbitrary. However, I believe any reasonable mapping of the quantities r_{ij} onto $P(i \mid A)/P(i \mid \bar{A})$ will preserve the qualitative results.

larger structures—i.e., the N(A,*i*). However, there are complications. One complication is that we have to keep separate the cases where the cue is actually relevant and the cases where it is not. That is, some of the cues in the environment will be just "noise" and have no relationship to our current information-retrieval needs. We assume that, on average, the environment provides X relevant cues and Y irrelevant cues. With this in mind, we break up the calculation of $P(i \mid A)$ into three parts corresponding to the three ways that i might appear as a cue if A is relevant:

1. There is the possibility that i is a potential cue for A. However, this does not mean that it will occur as a cue because the potential cues might exceed the X slots for relevant cues. We assume that if there are X relevant cues and T target structures, then each structure gets on average X/T cues. The probability that i is one of these cues should be its proportion relatedness to A times the mean number of cues per structure. In other words, the probability in this case is:

$N[A,i] * (X/T)$

2. There is the possibility that i is a cue for some other target structure than A. The average probability of this happening should be a product of the mean number of cues per target, (X/T), the average proportion relatedness between a term i and structures other than A, $(N[\text{-MDUL}/A,i])$, and the number of structures i could be cueing, $(T - 1)$. So the average probability in this case is:

$(T-1) * N[\bar{A},i] * (X/T)$

3. The third possibility is that i is not a relevant cue for any structure but is just noise. The probability of this happening is a function of the ratio of noise cues to total number of cues. Thus, this probability is:

Y/m

where m is the total number of possible cues—the sum of all terms (as distinct from structures whose sum is M) in the data base. Combining 1., 2., and 3. we get:

$$P(i \mid A) = N[A,i]\times(X/T) + (T-1)\times N[\bar{A},i]\times(X/T) + Y//m \qquad (8)$$

The calculation of $P(i \mid \bar{A})$ can be broken into two cases.

1. Cue i is one of the relevant cues. The probability of this being true is basically the product of the average relevance of cue i to structures other than A $(N[\bar{A},i])$ and the number of cues (X). Thus:
   $X * N[\bar{A},i]$
2. The cue is an irrelevant cue. In this case, as before, the probability is:
   Y/m

Combining 1. and 2. we get:

$$P(i \mid \bar{A})/ = N[A,i]\times X + Y/m \qquad (9)$$

Finally, we can derive the formula for cue strength:

$$P(i \mid A \;/\; P(i \mid \bar{A}) = 1 + (N[A,i] - N[\bar{A},i])/K \qquad (10)$$

where $K = T(N[\bar{A},i] + Y/Xm)$

A cue strength of greater than 1 increases the likelihood ratio of A, whereas a cue strength less than 1 decreases the likelihood ratio. Thus, the likelihood ratio of A will be increased just in case the relatedness of i and A is greater than average. In a typical data base like human memory, a cue will have above-average relatedness to a few structures and slightly below-average relatedness to most structures. The average relatedness should also be relatively close to zero.

This analysis of cue strength can be directly related to the priming literature if we make an identification between association norms and cue relatedness, r_{ij}. It predicts that words like *dog* will prime judgments of related words like *cat*. The r_{ij} values between such words will be high. Therefore, the N[A,*i*] values between one word, like *dog*, and memory structures involving another word, like the structure encoding the spelling of *cat*, will be high. Thus, access to these structures will be facilitated. The mathematics we have gone through should not be allowed to obscure the basic rationality of this prediction. The basic prediction is that information such as the spelling of *dog* is made more available in the presence of related words like *cat* because in fact the subject is more likely to need such information in these contexts.

This analysis also predicts inhibitory effects, something my own theory, ACT*, cannot do. Irrelevant cues will have slightly less than average N[A,*i*] values to the target structure and so lower the odds ratio that defines the structure's likelihood. That is, for instance, a prime like *lip* should make knowledge about *dog* less available because *lip* is seldom present when knowledge about dog is required.

Finally, this analysis can predict another startling result that ACT* failed to handle. This is the observation that one cannot seem to get second-order priming. DeGroot (1983—see also Balota & Lorch, 1986; Ratcliff & McKoon, 1987) used triplets of words like bull–cow–milk where there is a strong association between the first and second and between the second and third but not between the first and third. The first did not prime the third as would be predicted by a spreading activation model in which activation would spread from the first to the second and hence to the third. However, on the preceding analysis, the first and third terms would have low relatedness. This is in fact the rational thing to do: If milk is never processed in the presence of bull, one should not prime structures involving milk when bull appears.

Thus, we see that this rational approach predicts three additional properties of human memory—raising our a priori prediction rate to 6 out of 6. Again, to remind the reader, these predictions are being made on the assumption that probability correct and reaction times are monotonically related to the probability that a structure is needed. These need probabilities are the *p[A]* from Equation 1 or the $p(A \mid H_A \ \& \ Q)$ from Equation 2.

FACT RETRIEVAL AND FAN EFFECTS

Much of my experimental life has been spent studying how subjects retrieve sentences that they have learned (Anderson, 1983). My favorite manipulation has been one where I have varied the number of facts that a particular concept appeared in. The present relatedness analysis can be extended to apply to these experiments.

To show how the fan effect falls out of this relatedness analysis, I will work with a particularly idealized situation where we assume that each term in the data base has a fan of f (that is, it occurs in f structures), each structure has e elements in it, and no two terms co-occur in more than one structure. These uniformity assumptions facilitate mathematical analysis, but the same points come through in more complicated ways if we assume that the memory is not so uniformly organized with respect to fan.

The relatedness between any term and itself is 1. The relatedness between a term and another term that occurs in the same proposition is $1/(2f-1)$. With all other terms, the relatedness is 0. Thus, we have:

$r_{ij} = 1$ *if* $i = j$
$= 1/(2f-1)$ if i and j occur in the same structure
$= 0$ otherwise.

The average relatedness between a structure and a term that appears in it is $(2f + e - 2)/e(2f-1)$. The average relatedness between a structure and a term that occurs with one of the structure's terms in another structure is $1/e(2f-1)$. Thus,

$R[A,i] = (2f + e - 2)/e(2f - 1)$ if i occurs in A
$= 1/e(2f - 1)$ if i and some term j in A co-occur in another structure
$= 0$ otherwise

We can now calculate the quantities $N[A,i]$ and $N[\bar{A},i]$, which go directly into calculating the ratio $P(i/A)/P(i/\bar{A})$:

$$\frac{N[A,i]}{} = \frac{2f + e - 2}{f[f(e + 1) - 1]}$$

$$\frac{N[\bar{A},i]}{} = \frac{(f - 1)[f(e + 1) + e - 2]}{[M - 1)f[f(e + 1) - 1]}$$

where M is the number of structures in the data base. $N[A,i]$ varies from 1 when $f = 1$ to 0 when $f = \infty$. $N[\bar{A},i]$ varies from 0 when $f = 1$ to $1/M-1$ when $f = \infty$. Thus, the difference between the two decreases with increasing fan and so the ratio $P(i \mid A)/P(i \mid \bar{A})$ decreases.

Increasing the fan has another effect besides lowering the cue strength of a particular structure. It also increases the number of structures to which the cue has non-zero relatedness. This can have an effect in two situations. First, when the subjects are in a situation where they have to exhaust all related structures, they should be slowed down. This is the situation when a subject must reject a foil in a typical fact-retrieval experiment. Thus, we predict the fan effect for foils, although on a different basis than the fan effect for targets.

Also, subjects are sometimes in situations where they can respond on the basis of any of a number of structures. In this case, the higher the fan, the more relevant structures there are, and the faster they should respond. This is basically a base rate effect in a Bayesian framework. This is the situation in the experiments (Reder & Ross, 1983; Reder & Wible, 1984) that have allowed subjects to respond if they studied anything thematically consistent with the query. Here, a "negative fan effect" is found. The more things the subjects have studied that are consistent with the theme, the faster they are to respond.

There are many interactions involving the fan effect. However, to address interactions rather than main effects, I would have to define a mapping from the likelihood ratio to the dependent measures of reaction time and probability. This mapping is necessary so we can determine which things are going to be additive and which are not. As already announced, this is not the task of the present chapter, so I have to be content with picking up three more predictive successes from this a priori analysis to get my total up to 9 out of 9.

ENCODING SPECIFICITY

It is only reasonable that my tenth predictive success should be the encoding specificity phenomenon (Tulving & Thomson, 1973), which Tulving used to haunt my previous theories. Consider the basic demonstration: Subjects study "train–black" and are later asked to recognize black in the presence of train (the best condition), alone (the next best condition), or in the presence of white (the worst condition). This result follows directly from the present analysis of cue strength. If a cue like "train" is part of the target trace, it will have greater similarity ($N[A,i]$) to the target trace and so will increase the likelihood ratio for the target trace. If there was nothing studied with the target, the interitem sim-

ilarities r_{ij} will not be affected and so the cue will have a base similarity to the trace. If a different cue was present at study, the similarity between the test cue and trace will be reduced due to the normalization in the calculation of N[A,*i*] (Equation 6) and so the likelihood ratio for the target trace will be lowered.

RELATIONSHIP TO GILLUND & SHIFFRIN

It is worth commenting on the similarity between this proposal for cue combination and the SAM (Gilland & Shiffrin, 1984) model. Both models share the following components:

1. The idea that the cues combine multiplicatively (Equation 2 in this chapter).
2. The idea that the critical intervening variable is strength of associations between items and memory structures (Equation 5 in this chapter).
3. The idea that strength of associations should be normalized against competing associations (Equation 6 in this chapter).

The actual mathematics by which the two theories achieve their predictions are quite different, but it is hard to appreciate the significance of these differences. There is one fundamental conceptual difference, however. In the SAM model, the associations between the terms and the structures (N[A,*i*] in this chapter) are conceived of as reflecting the amount of time which A and *i* were together in the buffer. In the present model, they are based on the measure of co-occurrence relatedness, r_{ij}, taken from the information retrieval literature.

In typical information-retrieval systems, the concept of a time in the buffer together has no meaning and so one could not derive a measure like that in SAM. However, I suspect that even if it were available, it would be judged inadequate. One cannot count on having all relevant item-structure pairs co-occur together in the past to yield stable estimates of the probability of a structure given the item as a cue (or vice versa). Even in the scheme proposed here, it is something of a leap of faith to go from interitem relatedness to such a probability, but at least one is aggregating over more experiences and not insisting that the item and the structure be part of the same experience. Thus, rationally the model presented here is to be preferred.

CONCLUSIONS

At least ten of the basic phenomena that memory researchers have labored long to document can be predicted by the principle of rationality. I submit that this is a rather startling outcome. However, it should be acknowledged that the principle

of rationality cannot apply to memory (or any other human domain, including economic behavior) without a framing of what the problem is. Conceivably, one might frame the problem of human memory differently and come up with different predictions. I would be less than honest if I did not admit that my knowledge of human memory influenced my framing of the basic memory problem. However, I think the framing is quite reasonable.

One of the reasons why I have stuck with ordinal predictions and refrained from mapping this analysis onto interval predictions about time and probability is that I cannot see an equally plausible way of further framing the problem to make that mapping. Thus, if we are going to take this rational analysis of memory beyond these ten first-order predictions, the major agenda item is study of the task facing human memory—that is, the task that the system has evolved to handle. Perhaps we can gather evidence for some way of further specifying the memory task.

Although the present results are preliminary, they do support the hypothesis that we can predict the phenomena of human memory from the assumption that it operates rationally. What are we to make of this result if it continues to hold up under further analyses? One might take the attitude that the experimental study of memory is unnecessary because its behavior can be deduced from a priori premises. I do not think this is the correct conclusion. It ignores the fact that it is not certain a priori how to frame the problem faced by human memory so that we can propose a rational solution. The experimental research provides guidance here. Moreover, the rationality hypothesis, even with a framing, is just a scientific claim and still requires experimental test. However, I do think that the hypothesis throws an amazing light on the years of experimental research into human memory. It seems that these experiments may have been telling us that human memory was designed rationally.

ACKNOWLEDGMENTS

I should acknowledge Lynne Reder's invaluable contribution to this paper. If she had not got me reading and thinking about the work in information retrieval, I would never have discovered the framing of the memory problem that I have presented. Lynne also went through the manuscript with me to help assure that I got it right. I am also grateful for the comments of Bob Bjork, Gus Craik, and Roger Ratcliff. This research was supported by grant BNS 8705811 from the National Science Foundation.

REFERENCES

Anderson, J. R. (1978). Arguments concerning representations for mental imagery. *Psychological Review, 85,* 249–277.

Anderson, J. R. (1983). *The architecture of cognition.* Cambridge, MA: Harvard University Press.

Balota, D., & Lorch, R. (1986). Depth of automatic spreading activation: Mediated priming effects in pronunciation but not in lexical decision. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 12,* 336–345.

Bookstein, A., & Swanson, D. R. (1974). Probabilistic models for automatic indexing. *Journal of the ASIS, 25,* 312–318.

Bookstein, A., & Swanson, D. R. (1975). A decision theoretic foundation for indexing. *Journal of the ASIS, 26,* 45–50.

Burrell, Q. L. (1980). A simple stochastic model for library loans. *Journal of Documentation, 36,* 115–132.

Burrell, Q. L., & Cane, V. R. (1982). The analysis of library data. *Journal of the Royal Statistical Society, Series A*(145), 439–471.

Crowder, R. G. (1982). *The psychology of reading: An introduction.* New York: Oxford University Press.

DeGroot, A. (1983). The range of automatic spreading activation in word priming. *Journal of Verbal Learning and Verbal Behavior, 22,* 417–436.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91,* 1–67.

Kosslyn, S. M. (1980). *Image and mind.* Cambridge, MA: Harvard University Press.

Newell, A., & Rosenbloom, P. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1–55. Hillsdale, NJ: Lawrence Erlbaum Associates.

Pylyshyn, Z. W. (1981). The imagery debate: Analogue media versus tacit knowledge. *Psychological Review, 88,* 1–24.

Ratcliff, R. & McKoon, G. 1988). A retrieval theory of priming in memory. Psychological Review, *915,* 385–408.

Reder, L. M., & Ross, B. H. (1983). Integrated Knowledge in different tasks: The role of retrieval strategy on fan effects. *Journal of Experimental Psychology: Learning, memory, and cognition, 9,* 55–72.

Reder, L. M., & Wible, C. (1984). Strategy use in question answering: Memory strength and task constraints on fan effects. *Memory & Cognition, 12,* 411–419.

Salton, G., & McGill, M. J. (1983). *Introduction to modern information retrieval.* New York: McGraw-Hill.

Stritter, E. (1977). *File migration.* Unpublished doctoral dissertation, Stanford University.

Townsend, J. T. (1974). Issues and models concerning the processing of a finite number of inputs. In B. H. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition* (pp. 133–185). Hillsdale, NJ: Lawrence Erlbaum Associates.

Tulving, E. (1983). *Elements of episodic memory.* London: Oxford University Press.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80,* 352–373.

Wickelgren, W. A. (1973). The long and short of memory. *Psychological Bulletin, 80,* 425–438.

NOT A PAGE

III CLASSIFICATION SYSTEMS FOR MEMORY

BLANK